Estatística Inferencial

Prof. Wagner Hugo Bonat

Departamento de Estatística
Universidade Federal do Paraná

Conceitos preliminares

Inferência estatística

- **População** → distribuição de probabilidade.
- Intuição → Como que a v.a. deve se comportar na população.
- Variável → variável aleatória.
- Parâmetros da distribuição de probabilidade → **parâmetros populacionais**.
- Como a partir da amostra estimar os parâmetros populacionais?

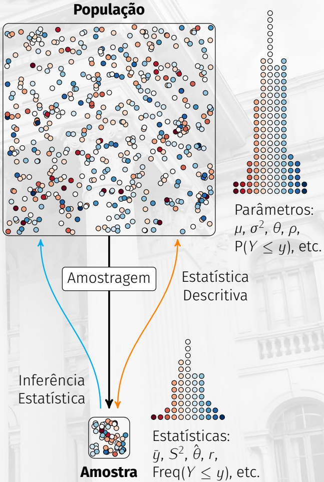

Figura 1. Processo de inferência estatística.

Inferência estatística

- Problema prático: Qual a proporção da população que desenvolveu anticorpos contra uma doença?
- Formalizando o problema:
  - Qual é a variável aleatória e quais valores ela pode assumir?
  - Y: desenvolveu anticorpos. Opções SIM ou NÃO.
- Qual a distribuição de probabilidade adequada para esta v.a.?
  - Bernoulli com função de probabilidade

$$\mathrm{P}(Y = y) = p^y(1-p)^{1-y}.$$

- Qual o parâmetro de interesse e o que ele significa?
  - p: proporção de pessoas que desenvolveram anticorpos.

Pensamento Estatístico

- Como determinar o valor de p?
  - Examinar todos os membros da população e verificar a proporção que desenvolveu anticorpos.
  - Examinar apenas alguns membros da população (amostra) e calcular a proporção que desenvolveu anticorpos.
- Problema: A proporção obtida na amostra não é a mesma obtida na população.
  - Incerteza associada ao valor da proporção devido a termos apenas uma amostra.
  - Como quantificar essa incerteza?
  - Como tomar uma decisão baseada apenas na amostra?
- Descrição probabilística da estatística de interesse $\rightarrow$ **Distribuição amostral.**

Especificação do problema de Inferência

- Y: desenvolveu anticorpos (v.a.).
- Especificação do modelo $Y \sim \text{Ber}(p)$.
- Parâmetro p.
- Informação sobre p através de uma amostra da população.
- Denotamos as amostras por $y_1, \ldots, y_n$.
- Objetivos da inferência estatística:
  - Estimar p baseado apenas na amostra (valor pontual)! Quanto é p na população?
  - Informar o quanto preciso ou creditável é o valor estimado (intervalo de confiança).
  - Decidir sobre possíveis valores de p baseado apenas na amostra.
  - A proporção da população com anticorpos atingiu um patamar desejável?

Especificação do problema de Inferência

- Suponha que coletamos uma amostra (aleatória) de tamanho $n = 10$ e que $y = 7$ pessoas apresentaram anticorpos.
- Qual valor você acha que o parâmetro p assume na população?
- Assumindo observações independentes, sabemos que a soma de v.a. Bernoulli é binomial com $n = 10$ e um parâmetro p desconhecido.
- Podemos calcular a probabilidade de observar $y = 7$ para um valor de p, por exemplo, $p = 0.8$

$$P(Y = 7|n = 10, p = 0{,}80) = \binom{10}{7} 0{,}80^7 (1 - 0{,}80)^{10-7} = 0{,}2013.$$

Especificação do problema de Inferência

- Para qualquer outro valor de p

$$P(Y = 7|n = 10, p) = \binom{10}{7} p^7 (1 - p)^{10-7},$$

variando p temos a **função de verossimilhança**

$$L(p) \equiv P(Y = 7|n = 10, p) = \binom{10}{7} p^7 (1 - p)^{10-7}.$$

- **Ideia**: Se p for um determinado valor, **qual a probabilidade** de observar o que eu realmente observei na amostra.

Pensamento frequentista

- Se o experimento for repetido um número grande de vezes e a cada realização $\hat{p}$ for obtido, o que aconteceria?
- $\hat{p}$ é uma variável aleatória.
- Se é variável aleatória, então tem distribuição de probabilidade que descreve o seu comportamento.
  - Qual é a sua distribuição?
  - Qual o seu valor esperado?
  - Qual a sua variância?
- Bom momento para revisar cálculo diferencial integral

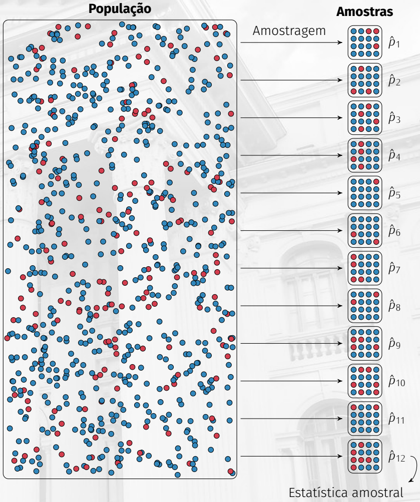

Figura 2. Ilustração da distribuição amostral.

Distribuição amostral

- Objeto de inferência (*frequentista*).
- A estimativa pontual é um resumo desta distribuição.
- Intervalos entre quantis representam a incerteza sobre o valor estimado.
- Compara-se estimadores concorrentes pelas características de suas distribuições amostrais.
- E para tudo isto:
  é preciso saber como estimar.

Figura 3. Distribuição amostral de diferentes estimadores de um parâmetro.

Componentes da modelagem estatística

- Modelo → comportamento da natureza.
- Parâmetros do modelo → parâmetros populacionais de interesse.
- Qual modelo melhor descreve os dados?
- Assumimos um modelo → parâmetros são desconhecidos.
- Baseado na amostra → encontrar os parâmetros compatíveis com a amostra.
- Descrever a incerteza → **distribuição amostral**.

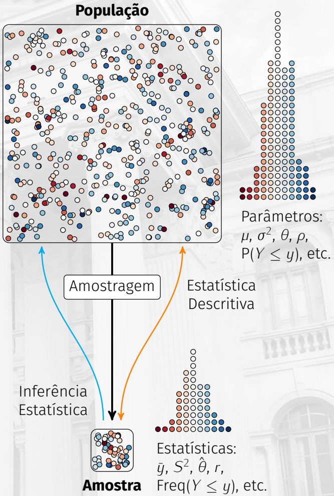

Figura 4. Processo de inferência estatística.

Estimadores e suas propriedades

Notação e definições (relembrando)

- $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$: v.a.'s independentes e idênticamente distribuídas.
- $Y_i \sim f(\boldsymbol{\theta})$ onde f denota a função densidade de probabilidade ou função de probabilidade e $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ é um vetor de p parâmetros populacionais.
- $\mathbf{y} = (y_1, \ldots, y_n)^\top$ denota o vetor de valores observados da v.a. Y.
- **Estatística** - Uma estatística é uma variável aleatória $\mathrm{T} = \mathrm{t}(\mathbf{Y})$, onde a função $\mathrm{t}(\cdot)$ não depende de $\boldsymbol{\theta}$.
- **Estimador** - Uma estatística T é um estimador para θ se o valor realizado $\mathrm{t} = \mathrm{t}(\boldsymbol{y})$ é usado como uma estimativa para o valor de θ.
- **Distribuição amostral** - A distribuição de probabilidade de T é chamada de distribuição amostral do estimador $\mathrm{t}(\mathbf{Y})$.
  - **Objetivo:** O que caracteriza **bons** estimadores? $\rightarrow$ Propriedades dos estimadores.

Questões

- O que torna um estimador “bom” em termos práticos?
- Existe “erro” na estimação? Como medir?
- Quais as propriedades desejáveis de um estimador?
- Como comparar dois (ou mais) estimadores?

Figura 5. Foto de cottonbro no Pexels.

Vício de um estimador

- Um estimador deve fornecer **valores próximos** do valor verdadeiro do parâmetro que está sendo estimado.
- Um estimador é **não viciado** para θ se o valor esperado de $\hat{\theta}$ for igual a θ.
- Isso quer dizer que a **média da distribuição amostral** de $\hat{\theta}$ é θ.
- Em certos casos, é possivel determinar o vício de um estimador de forma **analítica**.
- Em situações mais complexas, pode-se determinar de forma **computacional**.

Figura 6. Foto de icono.com no Pexels.

Vício de um estimador

- **Viés** - O viés de um estimador T é a quantidade

$$B(T) = E(T - \theta).$$

- O estimador T é dito não viciado para θ se $B(T) = 0$, tal que $E(T) = \theta$.
- O estimador T é assintóticamente não viciado para θ se $E(T) \rightarrow \theta$ quando $n \rightarrow \infty$.

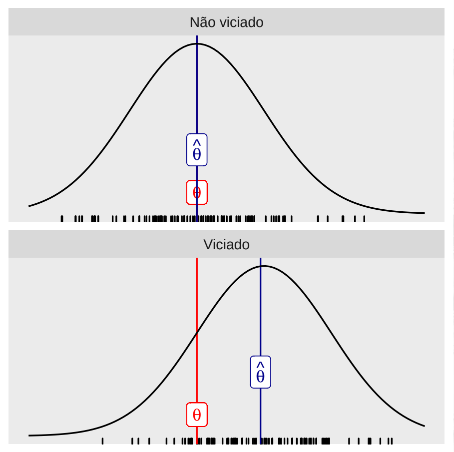

Figura 7. Exemplo de estimador viciado e não viciado.

Exemplo: Distribuição Normal

- Sejam $Y_i \sim N(\mu, \sigma^2)$, para $i = 1, \ldots, n$.
  a) Mostre que o estimador $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ é não viciado para μ.
  b) Mostre que o estimador $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$ é viciado para σ^2 e determine o seu viés.
  c) Mostre que o estimador $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$ é não viciado para σ^2.
  d) Compare os estimadores para σ^2 em b) e c) em termos assintóticos. O que você pode concluir?

Variância de um estimador

- Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ estimadores não viciados de θ.
- Então, $\text{E}(\hat{\theta}_1) = \text{E}(\hat{\theta}_2) = \theta$.
- No entanto, as **variâncias** destas distribuições amostrais podem ser diferentes.
- É razoável escolher o estimador que apresente a **menor variância**.

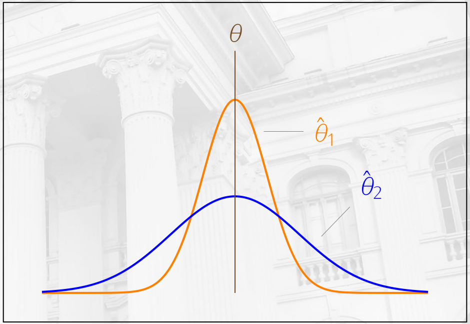

Figura 8. Distribuição amostral de dois estimadores não viciados.

Erro quadrático médio

- Nem sempre se dispõe de estimadores não viciados.
- Há situações em que estimadores viciados tem distribuição amostral com menor variância.
- Como escolher o estimador neste caso conciliando ambos aspectos, vício e variância?

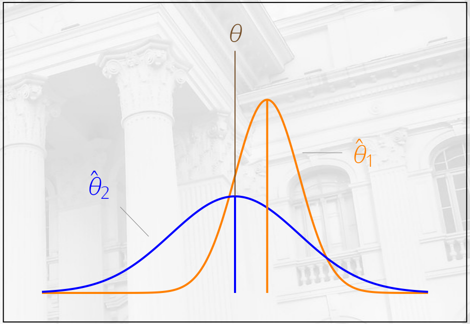

Figura 9. Distribuição amostral de dois estimadores.

Decomposição em vício e variância

- O **erro quadrático médio** (EQM) é uma medida que concilia vício e variância.
- O EQM de um estimador T do parâmetro θ é definido como

$$\text{EQM}(T) = E(T - \theta)^2.$$

- Ele pode ser reescrito como função da **variância** e **vício**

$$\begin{aligned} \text{EQM}(T) &= E[T - E(T)]^2 + [E(T) - \theta]^2 \\ &= V(T) + B^2(T). \end{aligned}$$

- Portanto, o EQM de um estimador não viciado é a própria variância.

Analogia do tiro ao alvo

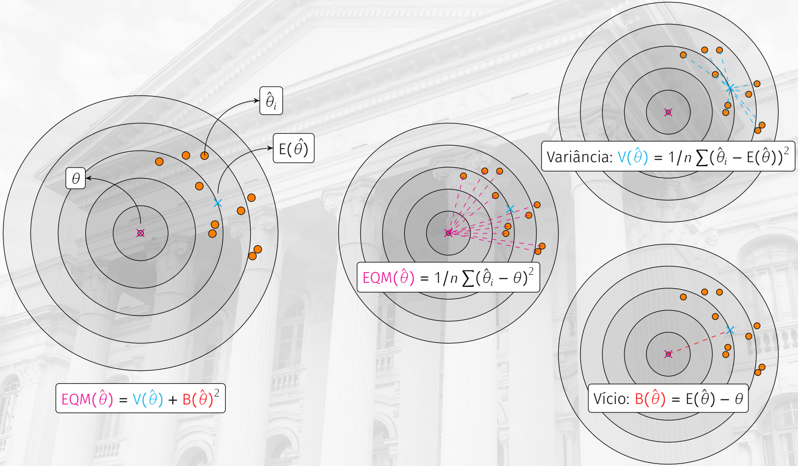

Figura 10. Analogia do tiro ao alvo para o erro quadrático médio e sua decomposição.

Eficiência relativa de um estimador

- O **erro quadrático médio** é uma métrica importante para comparar estimadores.
- Ele é usado para definir a **eficiência relativa** de um estimador comparado a outro,

$$\mathrm{Efr}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{EQM}(\hat{\theta}_1)}{\mathrm{EQM}(\hat{\theta}_2)}.$$

- Se a $\mathrm{Efr}(\hat{\theta}_1, \hat{\theta}_2) < 1$, conclui-se que $\hat{\theta}_1$ é um estimador superior à $\hat{\theta}_2$ e vice-versa.

Consistência de um estimador

- Não viés é uma propriedade desejável.
- Pode ser restrita em situações mais gerais.
- O viés de um estimador pode “sumir” quando a amostra aumenta de tamanho.
- Consistência é uma propriedade mais geral.
- Verifica o que acontece com o estimador quando a amostra aumenta de tamanho.

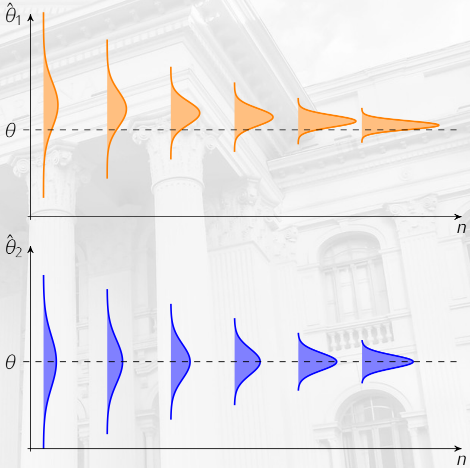

Figura 11. Consistência para dois estimadores.

Consistência de um estimador

- Verificar a consistência de um estimador não é trivial.
- Precisamos da idéia de **convergência** de v.a.
- Um estimador T é **médio quadrático consistente** para θ se o $\mathrm{EQM(T)} \rightarrow 0$ quando $n \rightarrow \infty$.
- O estimador T é **consistente em probabilidade** se $\forall \epsilon > 0$, $\mathrm{P}(|\mathrm{T} - \theta| > \epsilon) \rightarrow 0$, quando $n \rightarrow \infty$.
- Para consistência em probabilidade, a **Desigualdade de Chebyshev** permite dizer que

$$\mathrm{V}(\hat{\theta}) \rightarrow 0, \quad \text{para} \quad n \rightarrow \infty,$$

então $\hat{\theta}$ é consistente em probabilidade para θ.
- Existem outras formas de consistência $\rightarrow$ *Fisher consistency*.

Consistência do estimador $\hat{\sigma}^2$ da variância

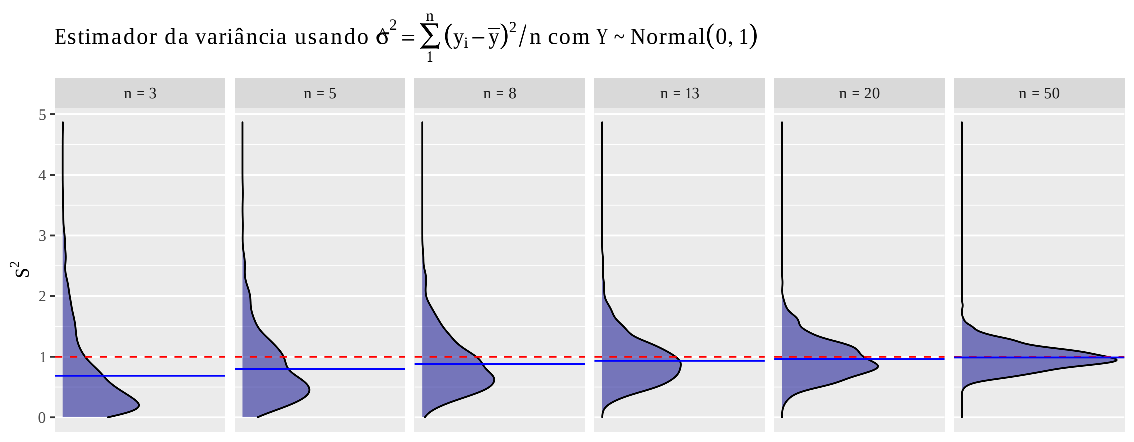

Figura 12. Ilustração por simulação computacional da consistência para o estimador $\hat{\sigma}^2$ da variância.

Inconsistência do estimador $\tilde{\sigma}$ do desvio-padrão

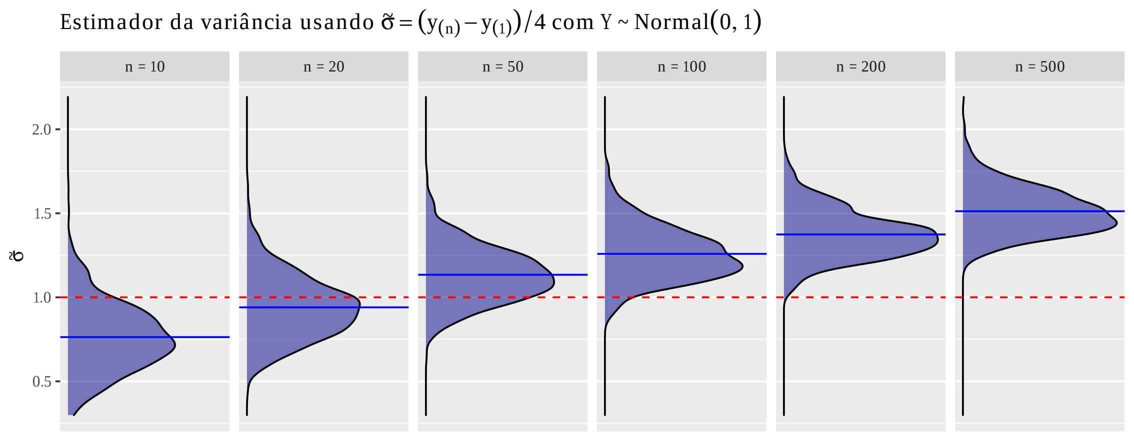

Figura 13. Ilustração por simulação computacional da inconsistência do estimador $\tilde{\sigma}$ do desvio-padrão baseado na regra empírica da amplitude.

Discussão

- O **estimador ideal** é aquele que captura a informação da amostra da forma mais eficiente.
- Deseja-se que seja não viciado, com a menor variância possível e consistente.
- A maioria dos estimadores vistos aqui apresentam tais características.
- Estimadores “empíricos” podem não apresentá-las.
- Há situações em que estimadores “óbvios” são superados por outros devidamente formulados.

Figura 14. Distribuição amostral de diferentes estimadores de um parâmetro.

Exemplo: Distribuição de Poisson

- Seja $Y_i \sim P(\lambda)$ para $i = 1, \ldots, n$ iid. Considere o estimador $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ para λ.
  a) Mostre que $\hat{\lambda}$ é não-viciado para λ.
  b) Encontre a variância de $\hat{\lambda}$.
  c) Encontre o erro quadrático médio de $\hat{\lambda}$.
  d) Mostre que $\hat{\lambda}$ é médio quadrático consistente.
  e) Mostre que $\hat{\lambda}$ é consistente em probabilidade.

Método da Máxima Verossimilhança

Função de verossimilhança

- **Função de verossimilhança** - Seja $\mathbf{y}$ um vetor $n \times 1$ representando uma realização de um vetor aleatório Y com função de probabilidade ou densidade probabilidade $f(Y, \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ denota um vetor $p \times 1$ de parâmetros, com $\boldsymbol{\theta} \in \Theta$, sendo Θ o respectivo espaço paramétrico. A função de verossimilhança ou simplesmente verossimilhança para $\boldsymbol{\theta}$ dado os valores observados $\mathbf{y}$ é a função $L(\boldsymbol{\theta}|\mathbf{y}) \equiv f(Y, \boldsymbol{\theta})$.
- Caso discreto:

$$L(\boldsymbol{\theta}|\mathbf{y}) \equiv P_{\boldsymbol{\theta}}(Y = \mathbf{y}).$$

- Intuição: Probabilidade de ver a amostra realizado dado que o parâmetro é um determinado valor.

Função de verossimilhança

- Caso contínuo: Probabilidade de um particular conjunto de valores ser observado é nula.
- Na prática medidas contínuas são tomadas com algum grau de precisão, digamos $y_{iI} \leq y_i \leq y_{iS}$. Assim,

$$L(\theta|\mathbf{y}) = P_\theta(y_{1I} \leq y_1 \leq y_{1S}, y_{2I} \leq y_2 \leq y_{2S}, \ldots, y_{nI} \leq y_n \leq y_{nS}).$$

- Esse é o caso mais geral e precisa da distribuição conjunta! Dificl de obter de forma geral.

Função de verossimilhança

- Supor que as observações são independentes, facilita

$$L(\boldsymbol{\theta}|\mathbf{y}) = P_{\boldsymbol{\theta}}(y_{1I} \leq y_1 \leq y_{1S}) \cdot P_{\boldsymbol{\theta}}(y_{2I} \leq y_2 \leq y_{2S}), \ldots, P_{\boldsymbol{\theta}}(y_{nI} \leq y_n \leq y_{nS}).$$

- Supondo ainda que todos os dados são medidos a um grau de precisão comum, digamos $(y_i - \delta/2 \leq Y_i \leq y_i + \delta/2)$. Assim, a verossimilhança fica

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= \prod_{i=1}^{n} P_{\boldsymbol{\theta}}(y_i - \delta/2 \leq Y_i \leq y_i + \delta/2) \\ &= \prod_{i=1}^{n} \int_{y_i-\delta/2}^{y_i+\delta/2} f(y_i, \boldsymbol{\theta}) dy_i. \end{aligned}$$

Função de verossimilhança

- Por fim, se o grau de precisão é alto e δ não depende dos valores dos parâmetros

$$L(\boldsymbol{\theta}|\mathbf{y}) \approx \prod_{i=1}^{n} f(y_i, \boldsymbol{\theta}).$$

- Resumindo, produto das distribuições marginais vista como uma função dos parâmetros.
- Caso geral, ainda precisamos de uma distribuição multivariada

$$L(\boldsymbol{\theta}|\mathbf{y}) \approx f(\mathbf{y}, \boldsymbol{\theta}).$$

Estimativas e estimadores

- **Estimativa de máxima verossimilhança** - Seja $L(\theta|y)$ a função de verossimilhança. O valor $\hat{\theta} = \hat{\theta}(y)$ é a estimativa de máxima verossimilhança para θ se $L(\hat{\theta}) \geq L(\theta)$, $\forall \theta \in \Theta$.

- **Estimador de máxima verossimilhança** - Se $\hat{\theta}(y)$ é a estimativa de máxima verossimilhança, então $\hat{\theta}(Y)$ é o estimador de máxima verossimilhança. Em geral vamos usar a abreviação EMV para nos referirmos ao estimador de máxima verossimilhança.

Verossimilhança: representações alternativas

- **Log-verossimilhança** - Se $L(\theta|y)$ é a função de verossimilhança, então $l(\theta|y) = \log L(\theta|y)$ é a função de log-verossimilhança.
- Segue do fato da função logaritmo ser monótona crescente que maximizar $L(\theta|y)$ e $l(\theta|y)$ levam ao mesmo ponto de máximo.
- **Verossimilhança relativa** - Sendo $L(\theta)$ a função de verossimilhança e sendo $\hat{\theta}$ a estimativa de máxima verossimilhança. A verossimilhança relativa é definida por $\frac{L(\theta)}{L(\hat{\theta})}$.
- **Função deviance** - Sendo $l(\theta)$ a função de log-verossimilha, a *deviance* é dada por $D(\theta) = -2[l(\theta) - l(\hat{\theta})]$.

Verossimilhança: representações alternativas

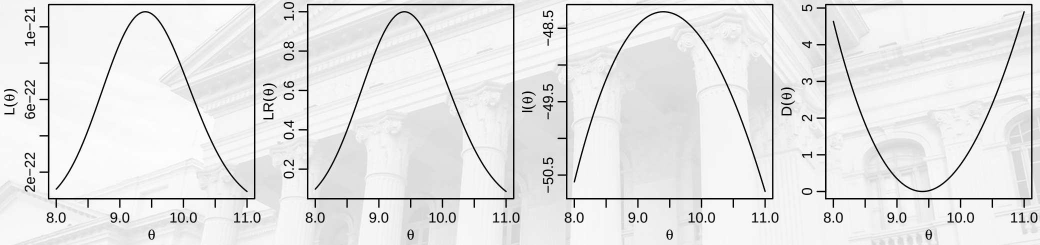

Figura 15. Diferentes formas de visualizar a função de verossimilhança.

Exemplo: Distribuição de Poisson

- Seja $Y_i \sim P(\lambda)$ para $i = 1, \ldots, n$ iid.
  a) Escreva a função de verossimilhança.
  b) Escreva a função de log-verossimilhança.
  c) Encontre o estimador de máxima verossimilhança.
  d) Escreva a função de verossimilhança relativa.
  e) Escreva a função *deviance*.
  f) Simule um conjunto de valores baseado neste problema e desenhe o gráfico das funções envolvidas (computacional).